MSc and Diploma in Applied Statistics 2010: Examiners' Report

Part I

A STATISTICS

(1) Numbers and percentages in each category

MSc in Applied Statistics

Category	**Number**				**Percentage**			
	2009/2010	2008/09	2007/08	2006/07	2009/10	2008/09	2007/08	2006/07
Distinction	6	8	5	6	18	28	17	13
Pass	19	18	25	35	58	62	83	78
Fail	7	2	0	1	21	7	0	2
Did not complete/ withdrawn	1	1	0	3	3	3	0	7

[Text removed from web version]

(2) Vivas

None

(3) Double marking of scripts

All scripts and dissertations were double marked.

B NEW EXAMINING METHODS AND PROCEDURES

There were no changes from recent years.

C CHANGES TO EXAMINING CONVENTIONS TO BE CONSIDERED

The change to the examination conventions made by the Supervisory Committee in Michaelmas Term 2009 (which was prompted by but different from the suggestions of the 2008-9 Examiners in their report) proved to be extremely helpful in considering candidates for distinction. The Examiners were happy for the revised 2009-10 conventions to be retained for 2010-11.

The Examiners asked the Supervisory Committee in June to request PDF copies of dissertations to ease the very busy scheduling of assessing and reviewing in the second half of September. This proved very useful and the Supervisory Committee has been asked to make this obligatory in future years.

The Examiners recommend to their successors that assessors for dissertations be asked not just for marks (for each of the criteria published in the course handbook) but also a brief (5-10 lines) report on each dissertation.

D EXAMINATION CONVENTIONS – COMMUNICATION WITH CANDIDATES

Draft conventions were set out in the course handbook distributed to students at the beginning of the academic year, with a revision issued by the Supervisory Committee at the end of Michaelmas Term 2009. A notice to candidates was issued in Hilary Term 2010 (attached), which included a copy of the revised conventions.

Part II

A GENERAL COMMENTS ON THE EXAMINATIONS

The internal examiners record their thanks to:

- The External Examiner and the assessors;
- Dr C N Laws for preparing the camera-ready papers
- Ms Jan Boylan and the other administrative staff of the Department of Statistics.

B EQUAL OPPORTUNITIES ISSUES AND BREAKDOWN OF THE RESULTS BY GENDER

Category	Male	Female
Distinction	5	1
Pass	6	13
Fail	2	5
Did not complete	0	1

C DETAILED NUMBERS ON CANDIDATES' PERFORMANCE IN EACH PART OF THE EXAMINATION

Paper I Principles of Statistical Analysis

Candidates were asked to attempt all 6 questions.

Question	Topic	No of answers	Median	IQR
1	Statistical Methods	33	11.0	4
2	Statistical Methods	33	10.0	6
3	Statistical Theory	33	8.0	5
4	Statistical Theory	33	10.0	6
5	Survival Analysis	33	10.0	4
6	Time Series	33	10.0	3

Paper II Further Statistical Methodology

Candidates were asked to attempt a total of 5 questions including 2 core questions.

Question	Topic	No of answers	Median	IQR
1	Computer Intensive Statistics	25	9.0	7
2	MCMC and Applied Bayesian Statistics	18	7.8	9
3	Further Statistical Methods	26	7.0	8.5
4	Infectious Diseases	13	11.0	10
5	Mathematical Genetics	6	12.0	2.3
6	Actuarial Science	15	9.0	3.5
7	Actuarial Science	23	7.0	9.0
8	Statistical Data Mining	10	10.5	7.5
9	Statistical Data Mining	6	17.0	3
10	Combinatorial Optimisation	21	12.0	7

D COMMENTS ON PAPERS AND INDIVIDUAL QUESTIONS

The Examiners made no overall comments.

Paper (i) Principles of Statistical Analysis

Q1 (Statistical Methods)

Part (a) and (b) were generally done well, but many students did not interpret appropriately the parameter lambda in the Box-Cox transformation. Part (c) was also done well, though students who tried to interpret the coefficient of pH in terms of Time (rather than log(Time)) generally did so incorrectly. Part (d.i) was not done very well, with many students saying there was a difference in causal interpretation between the two fits. Predictions were done well, but few students could describe how to obtain prediction intervals from R, and few appeared to notice that Time = 500 hours was well beyond the range of data used to fit the regression.

Q2 (Statistical Methods)

Part (a) was done reasonably well by almost all students. Throughout part (b) many students failed to state their answers in the context of the question. Most could answer the questions relating to hypothesis testing and interpretation of coefficients, though in several cases p-values were interpreted incorrectly. Few students could explain the relationship between the results of (b.iv) and (b.ii), and even fewer tackled (b.iii) sensibly.

Q3 (Statistical Theory)

Some students apparently did not know what the profile likelihood is. The calculations for the lognormal are easy if one has sufficient insight to recognize that, given the knowledge of alpha, the information is equivalent to that in a random normal sample. Not a single student thought to say that the likelihood and profile likelihood for the lognormal are zero if the minimum observation is less than alpha.

Q4 (Statistical Theory)

Some students misinterpreted 'associated Bayes estimator' (b.iii) thinking that this in itself is a defined term. Part (c) was answered best, although some of the answers contained little substance.

Q5 (Survival Analysis)

A straightforward and largely bookwork question that was not done well. Candidates managed the basic definitions, but e.g. few identified the use of a log-rank test.

Q6 (Time Series)

Done moderately well to well by almost all candidates. The relationship to least-squares regression (b.iv) foxed almost all.

Paper II

Q1 (Computer-Intensive Statistics)

A good test of the understanding of the material that was generally done well with some excellent answers.

The question stated clearly two methods of density estimation, kernel and logspline, but some candidates interpreted a histogram as a method of density estimation (which it is), and this was accepted in place of one of the others.

Q2 (MCMC and Applied Bayesian Statistics)

A relatively straightforward question that was done poorly except by two students, with the general level of understanding shown being very disappointing.

Q3 (Further Statistical Methods)

This question was not done well and a fair number of the answers were very incomplete. Some students seemed not to have read closely the text of the question, and not all were able to present the variances and covariances of dependent variables as implied by the model

specification. Some had difficulty in interpreting tables of parameter estimated, for example confusing standard errors of estimates with the estimated standard deviations of random effects.

Q4 (Infectious Diseases)

Some students wrote excellent answers. The attempts at parts (a-c) showed that many (but not all) had understood how to apply the SIR model to the new situation. The model in part (e) was quite tricky and only one candidate realised that the most obvious model contained an implausible assumption.

Q5 (Stochastic Models in Mathematical Genetics)

Candidates completed the bookwork parts well, but the unseen parts (b) and (c) were not answered satisfactorily.

Q6 (Actuarial Science)

Most students knew the symbols, but some did not express them in terms of the curtate lifetime K60. Several students wrote much more than the question asked in part (a.i). With a few exceptions, part (a.ii) was either poorly or not at all answered. Part (b) was answered well, but many students missed the opportunity to check whether their answers (based on the false conclusions in (a)) were realistic.

Q7 (Actuarial Science)

Parts (a) and (b) were bookwork and done satisfactorily. Part (c) was done well by the several students and most of those also had a good attempt at part (d). In part (d) several students constructed a new forward contract, while the question refers explicitly to Miss A, who was seen to enter 'the contract' three months previously.

Q8 (Statistical Data Mining)

The question was answered well in general. There were some issues in the definition of sensitivity and specificity in the context of ROC curves (even though in general full marks were awarded if the definitions were consistent). The definition of leave-one-out cross-validation and its relation to V-fold cross-validation could often have been stated more clearly.

Q9 (Statistical Data Mining)

Parts (a) and (b) were answered well, apart from numerical errors. The influence of bagging on the bias and variance of a regression procedure could often have been described better and the behaviour of the nearest neighbour method under bagging was often wrongly described.

Q10 (Combinatorial Optimization)

Part (a) asked for a solution ‘by Dijkstra's method’ and for path lengths to each node. Some students guessed a solution, and many gave the shortest path but not explicitly the lengths.

Attempts on parts (b) and (c) were generally good but some lacked clarity as to the exact steps being undertaken.

Dissertations

Dissertations were assessed on the following topics:

Dynamics of friendship networks and antisocial behaviour
Environments for health: understanding the correlates of stable housing among injection drug users
Wind power time series forecasting using statistical models
William Godwin - known by the company he keeps
Comparing the intensive care unit mortality predictions of random forests to those of logistic regression models
Inferring Partial Orders from Ranking Data
Modelling trends in mortality
Survival of children with *Rhabdomyosarcoma*
Pricing Consumer Derivatives on the Australian Football League Grand Final
Time series forecasting of volatility in stock market indices
On the identification and regulation of product failure
Statistical analysis of synchronous neural networks
Average quoted to average sold
Radiation-induced heart disease in patients with breast cancer
The competitive structure of the market (aggregator)
Understanding 200 years of legislative debate
Collaboration in teams: English Premier League
An analysis of copying price behaviour among insurance companies using multilevel model
Undergraduate Performance in Oxford colleges
Spatial biodiversity of tropical forest in Panama
Aggregations subjective ratings information
Statistics of packaging waste
Derivations and analysis of malaria parasite clearance times
Spatial models for treatment effect of statins on macrophage activity in carotid plaque inflammations
Document, implements and compare alternative forms of covariance matrix regularization in terms of their performance in portfolio allocation.
p univariate and multivariate density estimation for fire claims
Multiple imputation to estimate the effect of measurement uncertainty on the time course of sex ratio during gestation
Predicting one-year response in HIV clinical trials
Ancestral Inference from DNA Data
Effect of land use on successful pollination
Media contents and social network analysis in Uganda
Estimating within-group and between-group preservation bias - a maximum likelihood approach
Improving cooking stoves for India under clear development mechanism
A meta-analysis of heritable risk statistics for coronary heart disease
Analysis of objective polysomnography and subjective sleep measurements in insomnia patients
Analysis of proteomic data during and outside of a migraine attack

E. COMMENTS ON THE PERFORMANCE OF IDENTIFIABLE INDIVIDUALS

[Text removed from web version]

F EXAMINERS AND ASSESSORS

Examiners: Prof B.D Ripley (Chairman), Dr J Stander (External), Dr N Meinshausen, Professor T Snijders.

Assessors for examination papers: Professor R Griffiths, Professor C Holmes, Dr M Lunn, Professor C McDiarmid, Dr A Tomas, Dr M Winkel

Assessors for dissertations: Professor R Griffiths, Dr A D Lunn, Dr M Lunn, Dr C N Laws, Dr G Nicholls, Dr R Ripley, Dr D Steinsaltz, Dr A Tomas.